基于支持向量回归的生物测定数据分析

王志明[1,2,3]，谭显胜[1,2]，周　玮[1,2]，袁哲明[1,2,*]
（1. 湖南农业大学生物安全科学技术学院，长沙 410128；2. 湖南省作物种质创新与资源利用重点实验室，长沙 410128；
3. 湖南农业大学理学院，长沙 410128）

摘要：生物测定是生物学、医学、毒理学的重要内容与基础。常用的定量生物测定数据分析方法时间-剂量-死亡率模型（TDM）不能对复杂生测数据建立统一模型，信息利用不充分。本文基于支持向量回归（SVR），提出了一种能对不同供试因子、不同供试对象和不同环境条件下复杂生测数据统一建模的新方法。14 个简单生测数据和 2 套复杂生测数据的对比分析结果表明，SVR 模型拟合与留一法预测精度均优于 TDM 模型，估计的 LD_{50} 和 LT_{50} 等指标更为可信。SVR 模型有望作为 TDM 模型的有益补充，在定量生物测定数据分析中得到广泛应用。

关键词：时间-剂量-死亡率模型；互补重对数模型；支持向量回归；留一法；生物测定
中图分类号：Q965　　**文献标识码**：A　　**文章编号**：0454-6296(2010)12-1436-06

Bioassay data analysis based on support vector regression

WANG Zhi-Ming[1,2,3], TAN Xian-Sheng[1,2], ZHOU Wei[1,2], YUAN Zhe-Ming[1,2,*] (1. College of Bio-safety Science and Technology, Hunan Agricultural University, Changsha 410128, China; 2. Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha 410128, China; 3. College of Science, Hunan Agricultural University, Changsha 410128, China)

Abstract: Bioassay plays an important role in the studies of biology, medicine and toxicology. The time-dose-mortality model (TDM) widely applied to quantitative bioassay data analysis can not construct a unified model for complex bioassay data, and has the disadvantage of utilizing the information incompletely. Based on support vector regression (SVR), a novel quantitative bioassay model has been developed, which can construct a unified model for complex data with different test factors, different test objects and different environment factors. We compared the prediction performance between SVR and TDM using 14 simple data and 2 complex data. The results showed that SVR achieved better precision than TDM not only in self-consistency test but also in jackknife test, implying that the estimated values of LD_{50} and LT_{50} by the former are more reliable. As a useful supplement to TDM, SVR has the potential to be widely used for quantitative bioassay data analysis.

Key words: Time-dose-mortality model (TDM); complementary log-log model (CLL); support vector regression; Leave-One-Out method; bioassay

生物测定（bioassay）是生物学、医学、毒理学的重要内容与基础。早期定量生物测定数据的整理分析多采用 Finney 机值分析方法，但因其时间与剂量效应的评价相互脱节、不能揭示时间与剂量的互作效应、未充分利用实验信息等缺陷，近年来已逐步被时间-剂量-死亡率模型（time-dose-mortality model, TDM）所取代。TDM 按其数学结构又称为互补重对数模型（complementary log-log model, CLL），最早由 Preisler 和 Robertson 提出（Preisler and Robertson, 1989; Robertson and Preisler, 1992），并由冯明光等（1996）介绍引进国内，它综合考虑了时间与剂量效应并将其统一到同一个模型中，对多数简单生物测定数据具有较好的拟合能力，并方便给出 LD_{50} 和 LT_{50} 等各种指标，目前已广泛应用。

然而，TDM 模型仍然存在以下弊端：（1）CLL 是一个具显性表达式的非线性经验模型，很难假定所有生物测定数据均符合同一确定性模型而仅参数变化。对 CLL 模型拟合优度的评估常采用 Pearson

基金项目：湖南省杰出青年基金（10JJ1005）；湖南省教育厅青年基金（05B025）；湖南省教育厅科研项目（09C514）；湖南省高校科技创新团队项目

作者简介：王志明，男，1977 年生，湖南临澧人，博士研究生，副教授，生物信息和数据处理研究方向，E-mail：w3z2m1@163.com
* 通讯作者 Corresponding author, E-mail: zhmyuan@sina.com

收稿日期 Received: 2010-05-28；接受日期 Accepted: 2010-10-31

卡方检验或 Hosmer-Lemeshow 检验（唐启义和冯明光，2007）。在实际应用中，即使采用较不严格的 Hosmer-Lemeshow 检验，仍有部分生测数据不能通过。排除实验误差造成的影响，对这一事实的最可能解释是：这些生测数据本身不适于用 CLL 模型分析。（2）昆虫死亡与否属二项分布，对 CLL 模型参数的估计在 DPS 软件中是通过二项分布最大似然函数逼近来获得的，但实际应用中有些生测数据不符合二项分布（如杀菌剂或除草剂的时间-剂量-相对抑制率），限制了 CLL 模型的广泛应用（柏连阳等，2003）。最近，唐洁等（2006）采用非线性最小二乘法估计 CLL 模型参数，提出了杀菌剂生物测定动态评价时间-剂量-抑菌率模型，但该参数估计方法及由此获得的 LD_{50} 和 LT_{50} 等指标估计精度尚需更多研究实例的支持。（3）基于 CLL 的时间-剂量-死亡率模型与时间-剂量-抑菌率模型不能对不同供试因子、不同供试对象和不同环境下的复杂生物测定数据建立统一模型，导致实验信息利用不充分。例如分析在 m 种温度下 n 个球孢白僵菌 *Beauveria bassiana* 菌株不同剂量对桃蚜 *Myzus persicae* 不同时间段的死亡率，需构建 $m \times n$ 个独立的 CLL 模型。我们知道，统计分析中多样本的比较需采用方差分析，如对多样本采用 t 检验进行两两比较，不但繁琐、犯 α 错误的可能性增大，估计试验误差时的精度也有所损失。类似地，对一组复杂生物测定数据也应统一到同一个模型中，而不宜构建多个独立的 CLL 模型。为方便表述，本文称单一供试因子、单一供试对象和单一环境下的生物测定数据为简单数据，多种供试因子和/或多种供试对象和/或多种环境下的生物测定数据为复杂数据。（4）TDM 模型隶属于 Fisher 经典参数统计理论，把回归分析表达为特定参数化模型（CLL 模型）的参数估计问题；但经典参数统计理论是样本数目趋于无穷大时的渐近理论，它基于经验风险最小（训练误差最小），要求训练样本充分大。实际生物测定数据往往是小样本（这里样本数指的是剂量与时间段等形成的组合数，而不是每一剂量下的供试生物个体数）。在有限样本下，训练误差小并不总能获得好的预测效果，某些情况下，训练误差过小反而会导致推广能力的下降，即真实风险的增加，这就是过拟合或过学习问题（Vapnik，1995）。（5）生物测定数据分析的主要目标仅仅是求出未知函数在目标点的值，如 LD_{50} 和 LT_{50}。经典参数统计理论采用归纳和演绎分两步解决：第一步（从特殊到一般，即归纳），利用最大似然法从一个给定的函数集（CLL 模型）解决一个一般性问题；第二步（从一般到特殊，即演绎），从一般性问题出发，计算目标点的值。当拥有信息量（小样本）只够解决欲求问题，却不足以解决一个一般性问题时，用一个给定的函数集来直接估计目标点上的函数值，形成了一种新的从特殊到特殊的推理方法——转导推理（Vapnik，1995）。转导推理即小样本集推理的主要原则是：如果对欲求解的某一问题只拥有有限的信息，那么应该直接求解问题，而决不能求解一个更一般的问题来作为一个中间步骤，因为可能所拥有的信息足以直接求解问题，但不足以解决一个更一般的中间问题（邓乃扬和田英杰，2004）。

支持向量机（support vector machine，SVM）是统计学习理论的集大成者，它基于结构风险最小与转导推理，较好地解决了小样本、非线性、过拟合、维数灾难和局部极小等问题，泛化推广能力优异（Vapnik，1995；邓乃扬和田英杰，2004）。SVM 包括支持向量分类（support vector classification，SVC）和支持向量回归（support vector regression，SVR）。本文针对 TDM 模型存在的缺陷，建立了基于 SVR 的定量生物测定数据分析新方法，并以多个实例验证了新方法的有效性。

1　数据与方法

1.1　数据来源

14 个简单生物测定数据和 2 个复杂生物测定数据引自表 1 和 2 所列文献。

1.2　LIBSVM 2.8 软件包与 SVR

LIBSVM 2.8 软件包的 SVR 功能需用到 4 个常用程序：svmscale 用于对原始数据规格化，svmtrain 用于训练，svmpredict 用于预测，gridregression.py 用于自动搜索核函数最优参数 c，g，p（$c \in [-1, 6]$，$g \in [-8, 0]$，$p \in [-8, -1]$，步长均为 1）。各程序用法及其参数设置参见 Chang 和 Lin（2001）。

1.3　基于 SVR 的生测数据分析

1.3.1　拟合：以观察累积死亡率 y（%）为因变量，供试因子 x_1、供试对象 x_2、剂量 x_3、时间 x_4、环境因子（如温度）x_5 等为自变量构成二维数据表。其中供试因子 x_1 与供试对象 x_2 为单位向量。如供试对象为小菜蛾 *Plutella xylostella* 对 Bt 敏感品系与抗性品系，则敏感品系（$x_{2,1}$）可编码为 (0，1)，抗性

品系($x_{2,2}$)编码为 (1,0) 。一般地，复杂生测数据中供试因子和供试对象向量维数分别为供试因子数和供试对象类别数。简单生测数据自变量不包括供试因子、供试对象与环境因子，仅含剂量与时间。

数据集经 svmscale 规格化后，选用径向基核函数，以 gridregression. py 经格点搜索获得最优参数 c, g, p 组合。取最优参数组合，以 svmtrain 训练建模，以 svmpredict 预测数据集自身，获得拟合均方误差(root mean square error，*RMSE*) 与决定系数 (determination coefficient，R^2)，作为模型拟合优劣的评价标准。

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y(i,\ \text{true}) - y(i,\ \text{predict})]^2} \quad (1)$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}[y(i,\text{true}) - \bar{y}_{\text{true}}][y(i,\text{predict}) - \bar{y}_{\text{predict}}]}{\sqrt{\sum_{i=1}^{n}[y(i,\text{true}) - \bar{y}_{\text{true}}]^2\sum_{i=1}^{n}[y(i,\ \text{predict}) - \bar{y}_{\text{predict}}]^2}}\right)^2 \quad (2)$$

式中，$y(i,\ \text{true})$ 为第 i 个累积死亡率观察值，$y(i,\ \text{predict})$ 为第 i 个累积死亡率预测值，$\bar{y}_{\text{true}}$ 为真实值均值，$\bar{y}_{\text{predict}}$ 为预测值均值。

1.3.2 留一法预测：模型评估不仅希望拟合精度高，更重要的是泛化推广能力强。可采用留一法 (Leave-One-Out method，LOO) 检验 SVR 模型的独立预测能力：从 n 个样本中每次拿出一个样本作为测试集，以剩下的 $n-1$ 个样本构成训练集，依次经 gridregression. py，svmtrain 和 svmpredict 等完成预测。同上计算 *RMSE* 与 R^2 值。

1.3.3 估计 LD_{50} 和 LT_{50}：在给定供试因子、供试对象、环境因子和时间下求 LD_{50}，可产生一个仅剂量按小步长变化的测试集，以 1.3.1 训练好的 SVR 模型预测，预测累积死亡率最接近 0.5 所对应的剂量即 LD_{50}。类似可求 LT_{50}。

1.4 参比模型 TDM

TDM 模型的累积死亡率拟合值、LD_{50} 与 LT_{50} 由 DPS 软件直接给出(唐启义和冯明光，2007)，并据此算得拟合 *RMSE* 与 R^2 值。DPS 软件不允许原始数据单个缺失，因此基于 DPS 不能直接给出 TDM 模型留一法预测结果。根据 TDM 模型原理 (Robertson and Preisler，1992)，本文以 Matlab 语言重新编写了 TDM 留一法预测程序并在 Linux 上运行通过。经非缺失数据运算检验，结果与 DPS 一致。对一套复杂生测数据，先以多个独立的 CLL 模型分别实施留一法预测，最后合并预测结果，计算留一法预测 *RMSE* 与 R^2 值。

2 结果与分析

基于 SVR 的 14 个简单生测数据和 2 个复杂生测数据定量分析结果见表 1 和 2。从 *RMSE* 值与 R^2 值看，无论是简单生测数据还是复杂生测数据，也无论是模型拟合还是留一法预测，SVR 模型均优于 TDM 模型。

3 讨论

定量生物测定数据分析的目标是估计 LD_{50} 和 LT_{50} 等指标以评估供试因子对供试对象的毒力大小，其实质是独立预测。在小样本条件下，传统基于经验风险最小的回归模型的主要缺陷是往往模型拟合精度尚可，但独立预测精度明显下降。因此，评估模型更应注重其独立预测能力。本文涉及的 14 个简单生测数据和 2 套复杂生测数据是检索有完整原始数据文献后随机选取的，结果显示 SVR 模型无论是拟合还是留一法预测精度均优于 TDM 模型，因而有理由相信 SVR 模型估计的 LD_{50} 和 LT_{50} 等指标更为可靠。

TDM 模型对不同供试因子、不同供试对象和不同环境下的复杂生物测定数据需单独建模，因而实验信息利用不充分。例如，某球孢白僵菌对桃蚜在 20℃ 和 25℃ 下的生测数据分析，假定观测时间为处理后 1 - 10 d 共 10 个时间单位、球孢白僵菌对数浓度为 1 ~ 7 共 7 个浓度，假设 20℃ 时第 10 天的 LC_{50} 靠近对数浓度 5，TDM 建模估计 20℃ 的 LC_{50} 时采用 20℃ 时第 1 天与对数浓度 1 的观察数据，而不采用 25℃ 时第 10 天与对数浓度 5 的观察数据，直觉上就欠合理。相反，SVR 模型可对不同供试因子、不同供试对象和不同环境条件下复杂生测数据统一建模，其实验信息利用更为充分，且可估计未观测环境条件下 (如 22℃) 的 LD_{50}、LT_{50} 等指标。

SVR 定量生物测定分析模型的主要缺陷包括：(1) SVR 是非参数模型，而 TDM 是机理模型，后者模型中各个参数有明确的生物学意义。(2) SVR 没有考虑试验样本大小，也不能给出 LD_{50} 和 LT_{50} 等估计指标的置信区间。相反，TDM 建模时将每头试虫作为一个样本进行了考虑，且能给出 LD_{50} 和 LT_{50}

等估计指标的置信区间。（3）SVR 应用稍嫌复杂，运算速度偏慢。而 TDM 使用简便，运算速度快。现阶段，SVR 模型可作为定量生物测定数据分析中 TDM 模型的一个有益补充。

表 1　简单生测数据的 SVR 模型和 TDM 模型比较

Table 1　Comparison of SVR model and TDM model on simple bioassay data

供试因子 Factors	供试对象 Objects	SVR 模型 SVR model 拟合 Fitting *RMSE*（R^2）	SVR 模型 SVR model 留一法 LOO *RMSE*（R^2）	SVR LD_{50}	SVR LT_{50}	TDM 模型 TDM model 拟合 Fitting *RMSE*（R^2）	TDM 模型 TDM model 留一法 LOO *RMSE*（R^2）	TDM LD_{50}	TDM LT_{50}	参考文献 References
Ac	Oh	0.0263（0.9904）	0.0412（0.9761）	0.00	1.9	0.0693（0.9330）	0.0849（0.8978）	−0.51	1.5	Robertson and Preisler, 1992
Za	Mp	0.0447（0.9702）	0.0557（0.9536）	1.58	4.9	0.0520（0.9594	0.0574（0.9504）	1.53	4.6	刘彩玲和冯明光, 1998
Pn	Mp	0.0094（0.9988）	0.0140（0.9974）	0.65	3.8	0.0249（0.9916）	0.0271（0.9900）	0.60	3.9	李雪雁等, 2003
Za	Mp	0.0922（0.8768）	0.0964（0.8671）	1.47	4.6	0.0964（0.8593）	0.1034（0.8385）	1.60	4.5	鲍佳生和冯明光, 2006
Zr	Mp	0.0520（0.9706）	0.0583（0.9612）	1.48	2.7	0.0557（0.9653）	0.0663（0.9504）	1.48	2.6	许谦和冯明光, 1999
Pd（F95127）	Nl	0.0279（0.9675）	0.0312（0.9592）	1.28	6.9	0.0332（0.9528）	0.0346（0.9479）	1.46	7.5	Xu *et al*., 1999
Pd（F95129）	Nl	0.0138（0.9962）	0.0187（0.9932）	1.38	3.5	0.0490（0.9543）	0.0510（0.9485）	1.41	5.0	Xu *et al*., 1999
Pd	Mp	0.0277（0.9906）	0.0374（0.9829）	0.32	3.2	0.0447（0.9752）	0.0490（0.9708）	0.29	3.1	梁勇, 2003
Ct	Mp	0.0195（0.9928）	0.0346（0.9777）	1.84	4.0	0.0387（0.9712）	0.0458（0.9604）	1.80	4.5	李伟, 2002
Bb（F-263）	Tt	0.0181（0.9964）	0.0294（0.9900）	5.30	4.8	0.0458（0.9759）	0.0500（0.9708）	5.01	4.7	徐华潮, 2003
Bb（浙 B）	Tt	0.0361（0.9864）	0.0458（0.9785）	4.70	4.1	0.0500（0.9738）	0.0566（0.9667）	4.89	3.9	徐华潮, 2003
Bb（B12）	Tt	0.0193（0.9970）	0.0332（0.9912）	4.30	3.9	0.0361（0.9890）	0.0412（0.9860）	3.99	3.7	徐华潮, 2003
Bb（REBb01）	Tt	0.0447（0.9799）	0.0707（0.9495）	4.40	3.9	0.0806（0.9339）	0.0883（0.9203）	4.52	4.2	徐华潮, 2003
Bb（Bxs）	Tt	0.0230（0.9956）	0.0436（0.9843）	4.80	3.6	0.0447（0.9833）	0.0529（0.9773）	4.38	3.5	徐华潮, 2003

LD_{50}：仅给出最长观察时间间隔下的 LD_{50} 或 LC_{50} Only the LD_{50} or LC_{50} according to the maximum time interval was listed；LT_{50}：仅给出最高剂量或最高浓度下的 LT_{50} Only the LT_{50} according to the maximum dose or the highest concentration was listed；Ac：乙酰甲胺磷 Acephate；Za：安徽虫瘟霉 *Zoophthora anhuiensis*；Pn：诺氏虫疠霉 *Pandora nouryi*；Zr：根虫瘟霉 *Zoophthora radicans*；Pd：飞虱虫疠霉 *Pandora delphacis*；Ct：块耳霉 *Conidiobolus thromboides*；Bb：球孢白僵菌 *Beauveria bassiana*；Oh：*O. herbivorous*；Mp：桃蚜 *Myzus persicae*；Nl：褐飞虱 *Nilaparvata lugens*；Tt：竹梢凸唇斑蚜 *Takecallis taiwanus*. 下同 The same below.

表 2 复杂生测数据的 SVR 模型和 TDM 模型比较

Table 2 Comparison of SVR model and TDM model on complex bioassay data

供试因子 Factors	供试对象 Objects	SVR 模型 SVR model				TDM 模型 TDM model				参考文献 References
		拟合 Fitting *RMSE* (R^2)	留一法 LOO *RMSE* (R^2)	LD_{50}	LT_{50}	拟合 Fitting *RMSE* (R^2)	留一法 LOO *RMSE* (R^2)	LD_{50}	LT_{50}	
Pd (F95127)	Nl	0.0197 (0.9908)	0.0346 (0.9710)	1.38	7.3	0.0429 (0.9561)	0.0447 (0.9485)	1.46	7.5	Xu *et al.*, 1999
Pd (F95129)				1.48	3.5			1.41	5.0	
Bb (F-263)	Tt	0.0346 (0.9890)	0.0490 (0.9661)	5.30	4.7	0.0537 (0.9732)	0.0600 (0.9635)	5.01	4.7	徐华潮, 2003
Bb (浙 B)				4.70	4.1			4.88	3.9	
Bb (B12)				4.00	3.9			3.98	3.7	
Bb (REBb01)				4.20	3.8			4.52	4.2	
Bb (Bxs)				4.50	3.5			4.38	3.5	

参 考 文 献 (References)

Bai LY, Wu QM, Luo K, Tan XP, Yao WC, Zhou XM, 2003. Effect of pretilachlor mixed with 4 sulfonylurea herbicides on safety of rice seedling. *Journal of Hunan Agricultural University (Natural Sciences)*, 29(3): 220 - 222. [柏连阳, 吴秋明, 罗宽, 谭小平, 姚文昌, 周小毛, 2003. 丙草胺与 4 种磺酰脲类除草剂混用对稻苗安全性的影响. 湖南农业大学学报(自然科学版), 29(3): 220 - 222]

Bao JS, Feng MG, 2006. Broomcorn millet grain cultures of the obligate aphid pathogen *Zoophthora anhuiensis* (Entomophthorales) and their infectivity to *Myzus persicae* (Hemiptera: Aphididae). *Acta Entomologica Sinica*, 49(3): 393 - 398. [鲍佳生, 冯明光, 2006. 安徽虫瘟霉的黍米培养及其对桃蚜的侵染力. 昆虫学报, 49(3): 393 - 398]

Chang CC, Lin CJ, 2001. LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Deng NY, Tian YJ, 2004. Support Vector Machine - A New Method in Data Mining. Science Press, Beijing. 77 - 162, 224 - 272. [邓乃扬, 田英杰, 2004. 数据挖掘中的新方法——支持向量机. 北京: 科学出版社. 77 - 162, 224 - 272]

Feng MG, Tang QY, Hu GC, Huang SW, 1996. Susceptibility of seven species of aphids to a *Beauveria bassiana* isolate - analysis of time-dose-mortality model. *Journal of Basic Science and Engineering*, 4(1): 22 - 33. [冯明光, 唐启义, 胡国成, 黄世文, 1996. 球孢白僵菌对七种蚜虫的感染反应——时间-剂量-死亡率模型分析. 应用基础与工程科学学报, 4(1): 22 - 33]

Li W, 2002. Investigation of Entomophthorales Resource in Shandong and Studies on the Strain (Taian00927) of *Conidiobolus thromboides* Drechsler. MSc Thesis, Shandong Agricultural University, Tai'an, Shandong. 44. [李伟, 2002. 山东省虫霉目真菌资源调查和块耳霉菌株(Taian00927)的研究. 山东泰安: 山东农业大学硕士学位论文. 44]

Li XY, Yang Y, Xu WA, Li ZH, 2003. Bioassay for the virulence of *Pandora nouryi* Humber against *Myzus persicae*. *Pesticides*, 42(7): 20 - 22. [李雪雁, 杨勇, 许维岸, 李照会, 2003. 诺氏虫疠霉对桃蚜的毒力测定. 农药, 42(7): 20 - 22]

Liang Y, 2003. Cadaver-mimic Cultures of Homoptera-pathogenic Fungus *Pandora delphacis* Grown on Solid Substrates and Their Biological Features and Functions. PhD Dissertation, Zhejiang University, Hangzhou. 40. [梁勇, 2003. 同翅目昆虫专性病原真菌飞虱虫疠霉的固体培养法及其虫尸模拟培养物的生物学特性与功能. 杭州: 浙江大学博士学位论文. 40]

Liu CL, Feng MG, 1998. The time-dose effect of the entomophthoralean fungus, *Zoophthora anhuiensis* against the green peach aphid, *Myzus persicae*, in bioassay. *Mycosystema*, 17(4): 361 - 366. [刘彩玲, 冯明光, 1998. 安徽虫瘟霉对桃蚜的生物测定与时间-剂量效应分析. 菌物系统, 17(4): 361 - 366]

Preisler HK, Robertson JL, 1989. Analysis of time-dose-mortality data. *J. Econ. Entomol.*, 82(6): 1534 - 1542.

Robertson JL, Preisler HK, 1992. Pesticide Bioassays with Arthropods. CRC Press, Boca Raton. 127 pp.

Tang J, Tang QY, Sun CH, Shen AH, 2006. Dynamic evaluation model for time-dose-inhibition analysis for fungicide bioassays. *Plant Protection*, 32(1): 34-37. [唐洁, 唐启义, 孙传恒, 沈爱华, 2006. 杀菌剂生物测定动态评价时间-剂量-抑菌率模型. 植物保护, 32(1): 34-37]

Tang QY, Feng MG, 2007. DPS Data Processing System for Practical Statistics. Science Press, Beijing. [唐启义, 冯明光, 2007. 实用统计分析及其 DPS 数据处理系统. 北京: 科学出版社]

Vapnik VN, 1995. The Nature of Statistical Learning Theory. Springer Verlag Press, New York. 87-189.

Xu HC, 2003. Biological Characteristics of *Takecallis taiwanus* (Takahashis) and Virulence of *Beauveria bassiana* (Bals.) against *T. taiwanus*. MSc Thesis, Zhejiang University, Hangzhou. 33. [徐华潮, 2003. 竹梢凸唇斑蚜生物学特性及白僵菌对其毒力研究. 杭州: 浙江大学硕士学位论文. 33]

Xu JH, Feng MG, Xu Q, 1999. The virulence of the entomophthoralean fungus *Pandora delphacis* to the brown planthopper, *Nilaparvata lugens*. *Entomologia Sinica*, 6(3): 233-241.

Xu Q, Feng MG, 1999. Bioassay for the virulence of the entomophthoralean fungus *Zoophthora radicans* against the green peach aphid *Myzus percisae*. *Journal of Zhejiang Agricultural University*, 25(2): 151-154. [许谦, 冯明光, 1999. 生物测定根虫瘟霉对桃蚜的毒力. 浙江农业大学学报, 25(2): 151-154]

（责任编辑：袁德成）